HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2019 February 12.

Published in final edited form as:

Methods Mol Biol. 2017 ; 1551: 161–169. doi:10.1007/978-1-4939-6750-6_9.

Chromosome-Range Whole-Genome High-Throughput Experimental Haplotyping by Single-Chromosome Microdissection

Li Ma[1,2], **Wenzhi Li**[3,4], and **Qing Song**[5,6,7,8]

[1]4DGenome Inc, Atlanta, GA, USA. lma@msm.edu

[2]Cardiovascular Research Institute and Department of Medicine, Morehouse School of Medicine, 720 Westview Dr SW, Atlanta, GA, 30310, USA. lma@msm.edu

[3]Cardiovascular Research Institute and Department of Medicine, Morehouse School of Medicine, 720 Westview Dr SW, Atlanta, GA, 30310, USA.

[4]Center of Big Data and Bioinformatics, First Affiliated Hospital of Medicine School, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

[5]4DGenome Inc, Atlanta, GA, USA. qsong@msm.edu

[6]Cardiovascular Research Institute and Department of Medicine, Morehouse School of Medicine, 720 Westview Dr SW, Atlanta, GA, 30310, USA. qsong@msm.edu

[7]Center of Big Data and Bioinformatics, First Affiliated Hospital of Medicine School, Xi'an Jiaotong University, Xi'an, Shaanxi, China. qsong@msm.edu

[8]Research Wing D-203, 720 Westview Drive, Atlanta, GA, 30310, USA. qsong@msm.edu

Abstract

Haplotype is fundamental genetic information; it provides essential information for deciphering the functional and etiological roles of genetic variants. As haplotype information is closely related to the functional and etiological impact of genetic variants, it is widely anticipated that haplotype information will be extremely valuable in a wide spectra of applications, including academic research, clinical diagnosis of genetic disease and in the pharmaceutical industry. Haplotyping is essential for LD (linkage disequilibrium) mapping, functional studies on cis-interactions, big data imputation, association studies, population studies, and evolutionary studies. Unfortunately, current sequencing technologies and genotyping arrays do not routinely deliver this information for each individual, but yield only unphased genotypes. Here, we describe a high-throughput and cost-effective experimental protocol to obtain high-resolution chromosomal haplotypes of each individual diploid (including human) genome by the single-chromosome microdissection and sequencing approach.

Keywords

Experimental; Haplotype; Chromosome-length; Whole-genome; High-throughput

1 Introduction

A “haplotype” refers to a group of alleles inherited on a single chromosome (Fig. 1) [1,2], A large number of statistical and computational methods have been developed to reconstruct haplotypes from conventional unphased genotype data [3–7], These methods suffer from short-phasing distance, switch errors, and ambiguities [4, 8]. These uncertainties or ambiguities on haplotype configuration create complications in genetic analysis [9]. A number of experimental approaches have also been developed to determine haplotypes in recent years [10–22].

Chromosomal haplotypes will be essential for functional interpretation of genomes, especially for studying the impact of cis-interactions on gene expression (Fig. 2). Recent ENCODE data shows that only 5% of DNase I hypersensitive sites (DHSs) lie within 2.5 kb of transcriptional start sites (TSSs), the remaining 95 % of DHSs are positioned distally [23]. In fact, proximity in sequence data may be a poor predictor of interactions generally. Chromatin is highly packed and organized in the nucleus [24], and regulatory elements and genes that are far apart in terms of genomic sequence can be brought together in the nucleus via the formation of three-dimensional loops [25–28]. Only ~7% of these looping interactions are with the nearest gene. And *cis*-interactions appear to occur more often than expected by chance even at distances greater than 200 Mb (that is, intra-chromosomal contact probabilities at this distance are much greater than the average contact probability between different chromosomes [29]). Thus, chromosomal haplotypes will be necessary for identifying the causal genetic variants by genetic studies.

In this chapter, we will describe an experimental pipeline to obtain chromosomal haplotypes in a high-throughput manner (Fig. 3). Besides the features on high-accuracy, high-throughput, low-cost, and applicability to all types of genetic markers [13, 17, 30–32], a unique feature of this technology is the extremely long length (or full chromosome-length) of the recovered haplotype.

2 Materials

1. 1. A drop (100 μL) of fresh whole blood or living cells of human or any multiploid organisms *(see* **Note 1**).
2. 2. Gibco® PB-MAX™ Karyotyping Medium, stored at 4 °C.
3. 3. $_{L}$-Glutamine, stored at room temperature.
4. 4. Actinomycin D (20 μg/mL). Working solution: add 0.5 mL of acetone to the 10 mg vial containing Actinomycin-D powder. Mix with 100 mL of RPMI 1640 in limited light conditions. Distribute into labeled and dated polystyrene tubes in 3 mL aliquots under sterile conditions. Immediately place into −10 °C to −20 °C, store aliquots at −20 °C.
5. 5. Ethidium Bromide (EB) (1 mg/mL). Working solution: add 1 mL of stock solution (10 mg/mL) to 9 mL dH_20 in a light proof bottle with parafilm around

the lid (EB is light sensitive). This working solution will expire in 6 months or at expiration date of stock solution, whichever is sooner.

6. 6. Phytohemagglutinin-PHA, stored at 4 °C.
7. 7. Gentamicin, stored at 4 °C.
8. 8. Colcemid™ Solution in PBS (10 μg/mL), stored at 4 °C.
9. 9. 0.075 M KC1 hypotonic solution, shelf life 2 weeks at room temperature.
10. 10. Carnoys Fixative (Methanol:Acetic Acid, 3:1). This is to be made fresh immediately before needed and is to be tightly capped when not in use.
11. 11. UV-light sliceable foiled slide (Vashaw Scientific).
12. 12. Giemsa staining solution.
13. 13. 0.2-mL Leica collecting tube for microdissection (Leica Microsystems).
14. 14. A whole-genome amplification (WGA) kit, such as Sigma GenomePlex WGA4 kit.
15. 15. QIAquick PCR purification kit.
16. 16. A high-throughput sequencer or a high-throughput array genotyping platform.
17. 17. A regular desktop or laptop computer with CPU of 3.0 GHz and 8 GB RAM on a 32 or 64 Window XP/7/8 or an Ubuntu 12 LTS system.

3 Methods

3.1 Single-Chromosome Isolation

1. Collect about 100 μL of whole blood (anticoagulated by sodium heparin) into a conical 15-mL centrifuge tube containing 5 mL of prewarmed complete PB Max Karyotyping medium with fetal bovine serum (FBS), $_{L}$-glutamine, phytohemagglutinin (PHA), and gentamicin. Incubate at 37 °C for about 48 h *(see* **Note 2**).
2. Thaw actinomycin-D in 37 °C water bath. Add 200 μL of Act-D solution and 100 μL of EB working solution to each blood culture tube under sterile conditions in a hood. Mix gently by inverting tubes. Incubate cultures for 30 min at 37 °C.
3. Add 50 μL of colcemid to a final concentration of 0.083 μg/mL. Mix by inverting again gently and incubate at 37 °C for 30 min.
4. While tubes are incubating, make fresh fixative (3:1 methanohglacial acetic acid) and place in the freezer to chill. Also at this time, put 75 mM KC1 in the water bath to warm to 37 °C.
5. After incubation with EB and colcemid, centrifuge at 1000 rpm (135 ×*g)* for 10 min.

6. Aspirate all but 0.3 mL of supernatant, gently resuspend cell pellet. Add 37 °C prewarmed 75 mM KC1, vortex gently to assure KCl is mixed well with the pellet. Incubate at room temperature (25 °C) for 15 min.

7. Add four to five drops cold fixative, gently mix by inverting tubes, and centrifuge 1000 rpm for 10 min. Return fix to freezer. Repeat this step three times. Finally, resuspend cell pellet in 5 mL of cold fixative.

8. Drip the cells onto a UV-light sliceable foiled slide to spread chromosomes.

9. Briefly stain the chromosomes with Giemsa (1:20) for 10 min, dip the slide into dH_20, and air dry.

10. Put the UV-light sliceable foiled slide under a laser microdissection microscope (ASLMD; Leica) *(see* **Note 3**). Use a computer mouse to draw a circle on the computer monitor, and then the computer will direct a laser beam to cut a small foil containing target single-chromosomes (Fig. 4). The foil will be collected into a Leica collecting tube with 9 μL of dH_20 in it *(see* **Note 4**).

3.2 Whole-Genome Amplification (WGA)

11. The collected foil will be directly used in subsequent experiments without DNA extraction.

12. Amplify the single chromosomes with a whole-genome amplification kit following the manufacturer's protocol. The following steps are described for the WGA with the Sigma GenomePlex WGA4 kit.

13. Add 1 μL of the Lysis and Fragment buffer, and then incubate at 50 °C for 1 h.

14. Heat the sample to 99 °C for EXACTLY 4 min and chill on ice.

15. Add 2 μL of the Single Cell Library Preparation buffer and 1 μL of Library Stabilization solution. Incubate at 95 °C for 2 min.

16. Cool the sample on ice, consolidate the sample by centrifugation, and place on ice.

17. Add 1 μL of Library Preparation Enzyme, mix thoroughly, and centrifuge briefly.

18. Incubate with the following cycles: 16 °C for 20 min, 24 °C for 20 min, 37 °C for 20 min, and 75 °C for 5 min.

19. Incubate the samples at 95 °C for 3 min followed by 35 cycles of 94 °C for 30 s and 65 °C for 5 min.

20. Purify the amplified product with the QIAquick PCR purification kit.

3.3 Haplotyping by Single-Chromosome Sequencing

21. The amplified DNA will be ready for a whole-genome genotyping or next-generation sequencing following the manufacturer's User Guide.

22. If using next-generation sequencing, users need to map the reads and call SNP alleles, producing a VCF formatted file. If using genotyping arrays, SNP alleles will be directly output *(see* **Note 5**).

23. Read out the single-chromosomal sequence directly from the next-gen sequence data. The SNP alleles from the single-chromosome sequencing or genotyping arrays will be chromosomal haplotypes.

24. Input the experimental haplotype data into HiFi software to obtain high-resolution whole-genome haplotypes *(see* **Note 6**).

4 Notes

1. The blood cells can be directly subjected to cell culture. If a tissue specimen is used, please follow the experimental procedure for the primary cell culture. The living cells released from tissue specimens (such as by trypsin digestion) will be grown in the media and arrested at metaphase following the same protocol described above.

2. Either fresh blood or lymphoblastoid cells or any living cells can be used. Correspondingly, the medium should be switched to the growth medium for the corresponding cell type. For example, when lymphoblastoid cells are used, cells will be grown in RPMI 1640 containing 15% FBS and a mitogen (such as PHA) for 45 h, followed by the same experimental Steps from colcemid treatment to chromosome microdissection as described above.

3. Isolation of single chromosomes can be done with chromosome sorting or microfluidics or any other new device.

4. The higher volume of dH_20 in the collection tube, the higher chance to successfully collect the dropped foil, and also a higher usage and cost to use all subsequent reagents. So the users need to adjust the volume to ensure a successful collection of the microdissected foil. Direct visualization of the foil in the tube after each microdissection under the microscope is usually recommended.

5. When choosing the Illumina genotyping arrays, and the GenomeStudio software (Illumina part # 11207066) to call alleles, please choose the Forward model rather than Top/ Bottom calling.

6. HiFi software needs three input files, the low-resolution experimental haplotypes obtained in the procedure described above, an unphased genotype dataset, and a reference panel. The reference panels can be downloaded from the International HapMap Project database (phase 2 public release 22, phase 3 public draft release 1, and phase 2+3 February 2009 release 27), the 1000 Genomes Project database, and any other databases that contain the haplotype data for specific populations. For Windows user, HiFi requires three input files, haplotype. txt, genotype.txt, and refHaplotype.txt. The input files are named following the above examples. The names should match with them exactly. Run the software with double click

the software. For Linux user, HiFi requires three input files, haplotype.txt genotype.txt, and refHaplotype.txt. If the input files are named following the above examples. HiFi can be run as /HiFi; otherwise, the file names can be named by the user and provided in the haplotype, genotype, refHaplotype order. Run the software with the following Command example: HiFi haplotype_fileName genotype_fileName refHaplotype_file-Name. HiFi takes a fourth parameter called MAFSTEP, which is the changing step of minor allele frequency. Default value for MAFSTEP is 0.1. Its value can be set between 0 and 0.5. Command example: HiFi haplotype.txt genotype.txt refHaplotype.txt 0.01

Acknowledgment

This work was supported by US National Institutes of Health grants (R21HG006173, RC4MD005964, HL003676, RR014758, RR003034, GM74913, G12MD007602, U54MD007588), an American Heart Association grant (09GRNT2300003).

References

1. Slatkin M (2008) Linkage disequilibrium— understanding the evolutionary past and mapping the medical future. Nat Rev Genet 9:477–485 [PubMed: 18427557]
2. Liu N, Zhang K, Zhao H (2008) Haplotype- association analysis. Adv Genet 60:335–405 [PubMed: 18358327]
3. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084–1097 [PubMed: 17924348]
4. Browning SR, Browning BL (2011) Haplotype phasing: existing methods and new developments. Nat Rev Genet 12:703–714 [PubMed: 21921926]
5. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5, el000529
6. Li Y, Wilier CJ, Ding J, Scheet P, Abecasis GR (2010) Mach: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34:816–834 [PubMed: 21058334]
7. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. Nat Methods 9:179–181
8. Kukita Y, Miyatake K, Stokowski R, Hinds D, Higasa K, Wake N, Hirakawa T, Kato H, Matsuda T, Pant K, Cox D, Tahira T, Hayashi K (2005) Genome-wide definitive haplotypes determined using a collection of complete hydatidiform moles. Genome Res 15: 1511–1518 [PubMed: 16251461]
9. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, Olason PI, Ingason A, Steinberg S, Rafhar T, Sulem P, Mouy M, Jonsson F, Thorsteinsdottir U, Gudbjartsson DF, Stefansson H, Stefansson K (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet 40(9):1068–75 [PubMed: 19165921]
10. Fan HC, Wang J, Potanina A, Quake SR (2011) Whole-genome molecular haplotyping of single cells. Nat Biotechnol 29:51–57 [PubMed: 21170043]
11. Kirkness EF, Grindberg RV, Yee-Greenbaum J, Marshall CR, Scherer SW, Lasken RS, Venter JC (2013) Sequencing of isolated sperm cells for direct haplotyping of a human genome. Genome Res 23:826–832 [PubMed: 23282328]
12. Kitzman JO, Mackenzie AP, Adey A, Hiatt JB, Patwardhan RP, Sudmant PH, Ng SB, Alkan C, Qiu R, Eichler EE, Shendure J (2011) Haplotype-resolved genome sequencing of a Gujarati Indian individual. Nat Biotechnol 29:59–63 [PubMed: 21170042]
13. Rao W, Ma Y, Ma L, Zhao J, Li Q, Gu W, Zhang K, Bond VC, Song Q (2013) High-resolution whole-genome haplotyping using limited seed data. Nat Methods 10:6–7 [PubMed: 23269372]

14. Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, Kertesz M, Snyder M (2014) Whole-genome haplotyping using long reads and statistical methods. Nat Biotechnol 32:261–266 [PubMed: 24561555]
15. Selvaraj S, Dixson JR, Bansal V, Ren B (2013) Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. Nat Biotechnol 31:1111–1118 [PubMed: 24185094]
16. Suk EK, McEwen GK, Duitama J, Nowick K, Schulz S, Palczewski S, Schreiber S, Holloway DT, McLaughlin S, Peckham H, Lee C, Huebsch T, Hoehe MR (2011) A comprehensively molecular haplotype-resolved genome of a European individual. Genome Res 21: 1672–1685 [PubMed: 21813624]
17. Ma L, Xiao Y, Huang H, Wang Q, Rao W, Feng Y, Zhang K, Song Q (2010) Direct determination of molecular haplotypes by chromosome microdissection. Nat Methods 7:299–301 [PubMed: 20305652]
18. Yang H, Chen X, Wong WH (2011) Completely phased genome sequencing through chromosome sorting. Proc Natl Acad Sci U S A 108:12–17 [PubMed: 21169219]
19. Zhang K, Zhu J, Shendure J, Porreca GJ, Aach JD, Mitra RD, Church GM (2006) Long-range polony haplotyping of individual human chromosome molecules. Nat Genet 38:382–387 [PubMed: 16493423]
20. Ding C, Cantor CR (2003) Direct molecular haplotyping of long-range genomic DNA with Ml-PCR. Proc Natl Acad Sci U S A 100:7449–7453 [PubMed: 12802015]
21. Peters BA, Kermani BG, Sparks AB, Alferov O, Hong P, Alexeev A, Jiang Y, Dahl F, Tang YT, Haas J, Robasky K, Zaranek AW, Lee JH, Ball MP, Peterson JE, Perazich H, Yeung G, Liu J, Chen L, Kennemer MI, Pothuraju K, Konvicka K, Tsoupko-Sitnikov M, Pant KP, Ebert JC, Nilsen GB, Baccash J, Halpern AL, Church GM, Drmanac R (2012) Accurate whole- genome sequencing and haplotyping from 10 to 20 human cells. Nature 487:190–195 [PubMed: 22785314]
22. Xiao M, Wan E, Chu C, Hsueh WC, Cao Y, Kwok PY (2009) Direct determination of haplotypes from single DNA molecules. Nat Methods 6:199–201 [PubMed: 19198595]
23. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Lenhard B, Tcwari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA. The accessible chromatin landscape of the human genome. Nature. 2012;489:75–82. [PubMed: 22955617]
24. Dekker J (2006) The three ‘c’s of chromosome conformation capture: controls, controls, controls. Nat Methods 3:17–21 [PubMed: 16369547]
25. Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2:292–301 [PubMed: 11283701]
26. Dekker J (2008) Gene regulation in the third dimension. Science 319:1793–1794 [PubMed: 18369139]
27. Miele A, Dekker J (2008) Long-range chromosomal interactions and gene regulation. Mol BioSyst 4:1046–1057 [PubMed: 18931780]
28. Phillips JE, Corees VG (2009) Ctcf: master weaver of the genome. Cell 137:1194–1211 [PubMed: 19563753]
29. Lieberman-Aiden E, van Berkum NL, Wilhams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293 [PubMed: 19815776]
30. Li W, Fu G, Rao W, Xu W, Ma L, Guo S, Song Q (2015) Genomelaser: fast and accurate haplotyping from pedigree genotypes. Bioinformatics 31:3984–3987 [PubMed: 26286810]

31. Li W, Xu W, Fu G, Ma L, Richards J, Rao W, Bythwood T, Guo S, Song Q (2015) High- accuracy haplotype imputation using unphased genotype data as the references. Gene 572:279–284 [PubMed: 26232609]

32. Ma Y, Zhao J, Wong JS, Ma L, Li W, Fu G, Xu W, Zhang K, Kittles RA, Li Y, Song Q (2014) Accurate inference of local phased ancestry of modern admixed populations. Sei Rep 4:5800

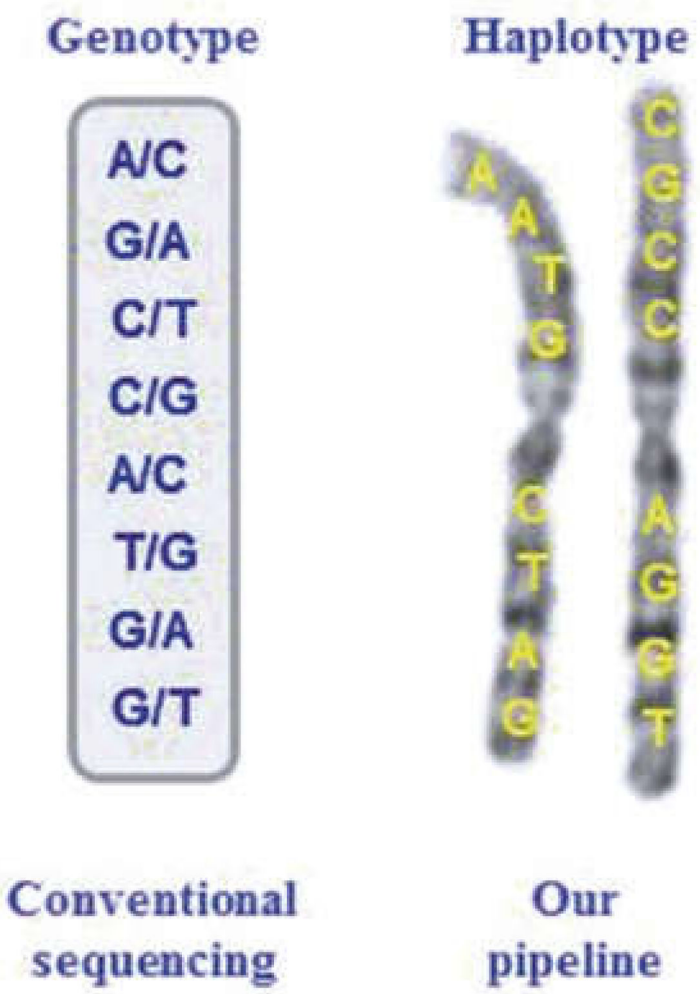

Fig. 1.
Difference between genotypes and haplotypes

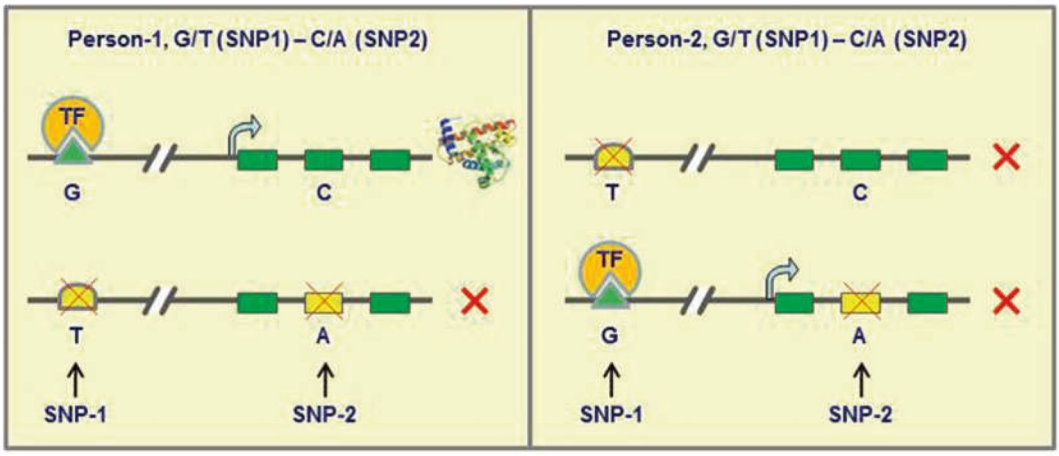

Fig. 2.

An illustration that haplotypes rather than genotypes are functionally relevant. In this special case, two different individuals have the same genotypes (G/T and C/A) but different haplotypes at two SNP loci. SNP-1 (G/T) occurred in an enhancer (G/T) that is essential for the expression of this gene, in which Allele-T is a null allele. SNP-2 (C/A) resides in an exon of this gene, in which Allele-A will disrupt the translation with an early stop codon. The production of this protein requires a cis-relatlonship between the enhancer and exon. Thus, Person-1 can produce this protein because one of his gene copies contains both functional alleles on the same chromosome (cis); Person-2 cannot produce this protein because neither of his gene copies contains two functional alleles. Please note that the enhancers may be close to a gene, or as far as 1 million base away from its regulatory target

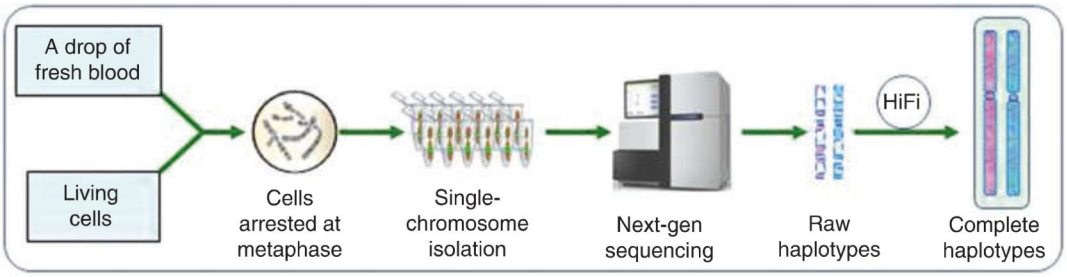

Fig. 3.

The pipeline of haplotype determination described in this chapter. In this pipeline, we will first culture the cells to metaphase and then isolate single-chromosomes; the isolated single-chromosomes will be then subjected to whole-genome amplification (WGA) followed by high-throughput sequencing; the sequencing data will be computationally analyzed (removing the noises, increasing the resolution and output the haplotypes)

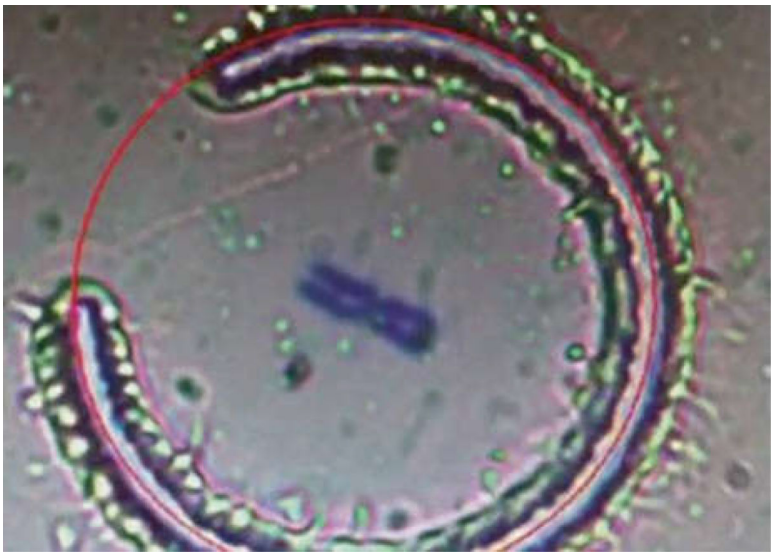

Fig. 4.
A chromosome is being microdissected by a laser beam directed by a computer